

The subject of this work is to propose a method for post-processing of scanned documents containing text and non-text elements. In this work, this problem is analyzed and divided into smaller sub-problems, which are solved. Main problems we propose solutions to are pre-processing of an input image to suppress scanning defects such as poor contrast or rotation of an input image, segmentation to characters, lines and images and finally compression of output document by clustering text elements. We will implement this proposed method as a program processing scanned images into PDF document.